

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 26, 2003, 11:15:29 ; Search time 12235 Seconds
(without alignments)
30.445 Million cell updates/sec

Title: US-10-086-184-1
Perfect score: 23
Sequence: 1 aatcgctccgagcgaggaac 23

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 60474

Minimum DB seq length: 0
Maximum DB seq length: 40

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estm:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12.6	54.8	39	10	AM248192 2819671.5
2	12.4	53.9	25	9	AA938266 0097601.8
3	12.4	53.9	34	9	AI354718 054412.x
4	12.2	53.0	28	9	AI646712 064407.x
5	12.2	53.0	37	9	AI966055 826604.y
6	12.2	53.0	37	17	TA381B01Q

Result	Score	Query Match	Length	DB ID	Description
7	12.2	53.0	40	9	AA939192 022209.8
8	11.8	51.3	37	9	AI474098 1A174098 tm05c07.x
9	11.8	51.3	39	17	A2430589 1M0215F10
10	11.6	50.4	38	17	TA137C040
11	11.6	50.4	39	17	A2800550
12	11.4	49.6	36	9	AA825458
13	11.4	49.6	37	14	H43693
14	11.2	48.7	27	9	AU254523
15	11.2	48.7	27	14	D18732
16	11.2	48.7	31	9	AA920864
17	11.2	48.7	37	13	B1561770
18	11.2	48.7	38	13	B1561718
19	10.8	47.0	22	9	AA985475
20	10.8	47.0	29	14	H55186
21	10.8	47.0	32	17	A2813220
22	10.8	47.0	33	17	A2803201
23	10.8	47.0	34	9	AA517484
24	10.8	47.0	35	2	HSN009944
25	10.8	47.0	35	17	A2469734
26	10.8	47.0	35	17	A2599526
27	10.8	47.0	36	12	BF533462
28	10.8	47.0	37	9	AI191467
29	10.8	47.0	37	9	AI735401
30	10.8	47.0	37	17	TA168H1P
31	10.6	46.1	34	10	BE386585
32	10.6	46.1	35	13	B1827700
33	10.6	46.1	37	9	AI538439
34	10.6	46.1	38	12	BF687876
35	10.6	46.1	39	17	BH866488
36	10.6	46.1	40	9	AI697005
37	10.4	45.2	26	17	BH759478
38	10.4	45.2	34	9	AA989545
39	10.4	45.2	34	10	AV833436
40	10.4	45.2	37	10	AV958431
41	10.4	45.2	37	17	TR21B11P
42	10.4	45.2	40	9	AI424339
43	10.4	45.2	40	17	BH857109
44	10.2	44.3	29	10	AA250348
45	10.2	44.3	32	12	BG718358

ALIGNMENTS

RESULT 1
LOCUS AM248192 39 bp mRNA linear EST 07-JUN-2000
DEFINITION 2819671.Sprine NIH_MGC_7 Homo sapiens cDNA clone IMAGE:2819671 5',
mRNA sequence.
ACCESSION AM248192
VERSION AM248192.1 GI:6591185
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 39)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE Unpublished (1999)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Other ESTs: 2819671.3prine
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DP CDNA Library Preparation: Ling
Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E.
Consortium (LNL) DNA Sequencing by: Berkeley MGC Sequencing
Project Clone distribution: MGC clone distribution information can
be found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnlnl.gov/bbrp/image/image.html Base Calling / Quality
Scores: PHRED from University of Washington Genome Center. Vector
Trimming: cross_match from University of Washington Genome Center
PHRAP suite. Poly-T Identification: patmatch.pl from Berkeley
Drosophila Genome Project. University of Washington Genome Center:

http://www.genome.washington.edu Low Quality Sequence: 0 contiguous
PHRED high quality bases following vector sequence. Very low
quality sequence. Trace file contained 39 contiguous distinct peaks
following vector sequence.

Plate: L1C2M2 row: D column: 8.

FEATURES

source

Location/Qualifiers
1. 39
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2819671"
/clone_lib="NIH MGC 7"
/issue_type="small cell carcinoma"
/cell_line="WGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Lung; Vector: pOTB1; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAG(G). Size selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the Laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT

7 a 11 c 17 g 4 t

ORIGIN

Query Match 54.8%; Score 12.6; DB 10; Length 39;
Best Local Similarity 78.9%; Pred. No. 1.5e+05;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAATCGGCTCGAGCGGCGG 19

Db 14 AAGCGCGCGCTGAGCGCGG 32

RESULT 2

AA938266

LOCUS

DEFINITION

AA938266 25 bp mRNA linear EST 27-AUG-1998
0097601.81 NCI CGAP Kid's Homo sapiens cDNA clone IMAGE:1574136 3'
similar to SW:CA12_MOUSE P28481 PROCOLLAGEN ALPHA 1(II) CHAIN
PRECURSOR ; mRNA sequence.

ACCESSION

AA938266 GI:3096377

VERSION

EST.

KEYWORDS

human.

SOURCE

Homo sapiens

REFERENCE

1 (bases 1 to 25)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher Moskalkuk, M.D., Ph.D., Michael R.

Emmett-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrip/image/image.html

Trace considered overall poor quality
Insert Length: 283 Std Error: 0.00
Seq primer: -40ml3 fwd. RT from Amersham
High quality sequence stop: 1.
Location/Qualifiers

FEATURES

source

1. 25
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1574136"
/clone_lib="NCI_CGAP_Kids"

/issue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AACTGAGAGATTCCGCGCCGCAATATTTTATTTTATTTT 3'],
double-stranded cDNA was ligated to Eco RI adapters
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

7 a 8 c 10 g 0 t

ORIGIN

Query Match 53.9%; Score 12.4; DB 9; Length 25;
Best Local Similarity 72.7%; Pred. No. 1.7e+05;
Matches 16; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 AAATCGGCTCGAGCGCGGAA 22

Db 4 AAGCGCGCGCGCGCGGAGAA 25

RESULT 3

A1354718

LOCUS

DEFINITION

A1354718 34 bp mRNA linear EST 04-JAN-1999
qt54h12.x1 Soares fetal_lung NBHL19W Homo sapiens cDNA clone
IMAGE:1951847 3' similar to SW:PRPB_HUMAN P2814 PROLINE-RICH
PEPTIDE P-B.; contains element MSRI repetitive element ; mRNA
sequence.

ACCESSION

A1354718 GI:4094871

VERSION

EST.

KEYWORDS

human.

SOURCE

Homo sapiens

REFERENCE

1 (bases 1 to 34)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES

High quality sequence stop: 1.
Location/Qualifiers

source

1. 34
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1951847"
/clone_lib="Soares fetal_lung_NBHL19W"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Lung; Vector: pT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
-TGTTACCATCTGAGTGGAGCGCGCCCAATTTTATTTTATTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NBHL19W."

BASE COUNT

8 a 7 c 15 g 4 t

ORIGIN

Query Match 53.9%; Score 12.4; DB 9; Length 34;

	Best Local Similarity	72.7%	Pred. No. 1.7e+05;
	Matches	16;	Conservative
		0;	Mismatches
		6;	Indels
		0;	Gaps
QY	1 AAATCGGCTCCGAGGCGGGAAA	22	
D8	1 AAAAGGGTTCCGGGGGGGGGAA	22	

RESULT 4	
A1646712	
LOCUS	28 bp mRNA linear EST 29-APR-1996
DEFINITION	ubc4h07.xl Soares mammary gland NMLMG Mus musculus cDNA clone IMAGE:138257.3; similar to TR:Q08805 Q08805 SALIVARY PROLINE-RICH PROTEIN L ; , mRNA sequence.

ACCESSION	A1646712
VERSION	A1646712.1
KEYWORDS	EST.
SOURCE	house mouse.

ORGANISM
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
1 (baes 1 to 28)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

UNPUBLISHED (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LINL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:905025

This clone was previously sequenced on the 5' end only, this new data is from the 3' end
Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand
High quality sequence: scop: 1.
Location/Qualifiers
1..28
source

```

BASE COUNT
ORIGIN
10 a          9 c          8 g          1 t
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      /clone="IMAGE:1382557"
      /clone_1lb="Soares_mammary_gland_NTLMG"
      /sex="Female (lactating)"
      /tissue_type="mammary gland"
      /lab_host="DH10B"
      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo. "

```

Query Match	53.0%	Score 12.2;	DB 9;	Length 28;
Best Local Similarity	82.4%;	Pred. No.2,1e+05;		
Matches 14;	Conservative	0;	Mismatches 3;	Indels 0;
				Gaps 0
QY	6	GGCTCCGAGCGGAGAA	22	
db	5	GGCCCCCAGGGGGAGAA	21	

RESULT 5	
A1966055/c	
LOCUS	A1966055
DEFINITION	37 bp mRNA linear EST 30-NOV-200
	gc26d04.y1 Gm-cl013 glycine max cDNA clone GENOME SYSTEMS CLONE ID
	Gm-cl013-1760 5' similar to SW:PSAD CUCSA P32869 PHOTOSYSTEM I
	REACTION CENTRE SUBUNIT II PRECURSOR ; , mRNA sequence.
ACCESSION	A1966055
VERSION	A1966055.1 GI:5760692

KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	Labbers R. to 3/77			
	Shoemaker, R., Keim, P., Vodka, L., Expelding, J., Corryell, V., Khanna			
	Bolla, B., Marra, M., Hillier, L., Kucab, T., Martin, J., Beck, C.,			
	Wyle, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers			
	, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schuck			
	, R., Ritter, E., Kohn, S., Shih, T., Jackson, Y., Cardenas, M., McCann			
	, R., Waterston, R. and Wilson, R.			
	Public Soybean EST Project			
	Unpublished (1999)			
	Contact: Shoemaker R/Public Soybean EST Project			

Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: east@watson.wustl.edu

Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand This clone is
available through: Resgen, Invitrogen Corp, 2130 South Memorial
Parkway Huntsville, AL 35801 For further information call: (800
) 533-4363 or contact via email: cu@resgen.com
Seq primer: -40RP from Gibco
High quality sequence stop: 1.

```

FEATURES
SOURCE
1..37
Location/Qualifiers
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1013-1760"
/clone_11b="Gm-c1013"
/tissue type="Whole seedlings, 2-3 week old seedlings, greenhouse grown"
/lab host="XU10-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2: XhoI; This cDNA library was constructed from mRNA isolated from whole seedlings of 2-3 week old greenhouse grown plants. The cDNA library was prepared using the Stratagene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XU10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Expelding."
7 a 18 c 2 g 10 c
BASE COUNT
ORIGIN

```

Query Match	53.0%	Score 12.2	DB 9	Length 37
Best Local Similarity	82.4%	Pred. No. 2.2e+05		
Matches 14	Conservative	0	Mismatches 3	Indels 0
QY 6 GACCTCCGAGGCGCGGAAA	22			
DB 36 GACCTTCGAGGCGGAAA	20			

RESULT 6	
TA381B01Q/c	
LOCUS	TA381B01Q
DEFINITION	37 bp DNA linear GSS 13-DEC-2000
	T. brucei sheared genomic DNA clone 381b01, reverse sequence,
	genomic survey sequence.
ACCESSION	AL497653
VERSION	AL497653.1 GI:11873375
KEYWORDS	GSS.

SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
REFERENCE 1 (bases 1 to 37)
AUTHORS Hall N., Bowman S., Lennard N.J., Doggett J., Atkin R., Chillingworth C., Ormond D., Harris B., El-Sayed N., Hou L., Melville S.E., Rajandream M.A. and Barrell B.G.
TITLE Direct Substitution
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nhleanger@ac.uk

COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999). Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES
source
1..37
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="381b01"

BASE COUNT 5 a 22 c 9 g 1 t

ORIGIN

Query Match 53.0%; Score 12.2; DB 17; Length 37;
Best Local Similarity 82.4%; Pred. No. 2.2e+05;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 CGGCTCCGAGCGGGGAA 21
|||||
Db 25 CGGCGCCGCGCGGCA 9

RESULT 7
AA939192/c 40 bp mRNA linear EST 01-MAY-1998
LOCUS og2209.g1 NCI CGAP GC4 Homo sapiens cDNA clone IMAGE:1587137.3
DEFINITION similar to SM.PPPE HUMAN P02811 BASIC PROLINE-RICH PEPTIDE P-E
AA939192
;contains KKR.b2 MER30 repetitive element;; mRNA sequence.
AA939192 GI:3099105
EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 40)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:
www-bio.linn.gov/bdrp/image/image.html

JOURNAL COMMENT
Trace considered overall poor quality

Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.
Location/Qualifiers
1..40
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1587137"
/clone_1ib="NCI CGAP GC4"
/tissue_type="pooled_germ cell tumors"
/lab_host="DH10B"
/note="Vector: pRTT3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from 3 pooled germ cell tumors, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRTT3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 6 a 23 c 6 g 5 t

ORIGIN

Query Match 53.0%; Score 12.2; DB 9; Length 40;
Best Local Similarity 82.4%; Pred. No. 2.2e+05;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 GGCTCCGAGCGGGGAA 22
|||||
Db 32 GGCCCGCGGGGAA 16

RESULT 8
A1474098 37 bp mRNA linear EST 14-APR-1999
LOCUS tm05c07.x1 NCI CGAP Col4 Homo sapiens cDNA clone IMAGE:215562.3
DEFINITION similar to TR:Q39865 Q39865 HYDROXYPROLINE-RICH GLYCOPROTEIN
A1474098
;contains element MSRI repetitive element;; mRNA sequence.
A1474098 GI:4327143
EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 37)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:
www-bio.linn.gov/bdrp/image/image.html

JOURNAL COMMENT
Trace considered overall poor quality
Insert Length: 968 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
Location/Qualifiers
1..37
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:215562"
/clone_1ib="NCI CGAP Col4"
/tissue_type="moderately-differentiated adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: PCMW-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT."

Average insert size 1.7 kb. Life Technologies catalog #: 11531-019"

BASE COUNT 6 a 9 c 21 g 1 t

ORIGIN

Query Match 51.3%; Score 11.8; DB 9; Length 37;
 Best Local Similarity 69.6%; Pred. No. 3.2e+05;
 Matches 16; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 AAATGGCTCCGAGCGCGGAAC 23
 Db 1 AAAGGGCCCCGGGGGCGCAAC 23

RESULT 9
 AZ430589 39 bp DNA linear GSS 03-OCT-2000
 LOCUS 1M0215F10F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 DEFINITION clone UUGC1M0215F10 F, DNA sequence.
 ACCESSION AZ430589
 VERSION AZ430589.1 GI:10554602
 KEYWORDS GSS
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 39)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Haml, C., Islem, H., Longacre, S., Mahmud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A. and Wright, D., Weis, R.
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 Unpublished (2000)
 CONTACT: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymer Research Bldg., 20 S. 2030 E., SLG, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: dunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 plate: 0215 row: F column: 10
 Seq primer: CGTTGTAAACGACGCGCAGT
 Class: plasmid ends
 High quality sequence stop: 39.
 Location/Qualifiers
 1..39
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0215F10"
 /clone_1lb="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD42ny; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptor complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 1 a 17 c 15 g 6 t

ORIGIN

Query Match 51.3%; Score 11.8; DB 17; Length 39;
 Best Local Similarity 86.7%; Pred. No. 3.2e+05;
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 CGGCTCCGAGCGCGG 19
 Db 13 CGCTCCAGCGCGG 27

RESULT 10
 TA137C04Q 38 bp DNA linear GSS 13-DEC-2000
 LOCUS T. brucei sheared genomic DNA clone 137C04, reverse sequence.
 DEFINITION genomic survey sequence.
 ACCESSION AL466262
 VERSION AL466262.1 GI:11835617
 KEYWORDS GSS
 SOURCE Trypanosoma brucei.
 ORGANISM Trypanosoma brucei
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 1 (bases 1 to 38)
 Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S.E., Rajandream, M.A. and Barrell, B.G.
 Direct Submission
 Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh@sanger.ac.uk
 Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).
 Email: nh@sanger.ac.uk
 Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES
 source 1..38
 /organism="Trypanosoma brucei"
 /strain="TREU927"
 /db_xref="taxon:5691"
 /clone="137C04"

BASE COUNT 9 a 11 c 14 g 4 t

ORIGIN

Query Match 50.4%; Score 11.6; DB 17; Length 38;
 Best Local Similarity 77.8%; Pred. No. 3.9e+05;
 Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 CGGCTCCGAGCGCGA 22
 Db 18 CAGCAACGAGCGCGGTA 35

RESULT 11
 AZ800550/c 39 bp DNA linear GSS 16-FEB-2001
 LOCUS 2M0058B10R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 DEFINITION clone UUGC2M0058B10 R, DNA sequence.
 ACCESSION AZ800550
 VERSION AZ800550.1 GI:12952808
 KEYWORDS GSS.

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
SOURCE
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches
QY
Db
RESULT 12
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
house mouse.
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eulestostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
1 (bases 1 to 39)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0058 row: E column: 10
Seq primer: CACACGAGAAACAGTACGACC
Class: plasmid ends
High quality sequence stop: 39.
Location/Qualifiers
1..39
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U96C2M0058E10"
/clone_lib="Mouse 10kb plasmid U96C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g14732114[g]AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

[illegible]

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@atcsom.mucfl.edu

Insert Size: 851
High quality sequence starts: 1
High quality sequence stops: 1
Source: IMAGE Consortium, LML
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.lml.gov) for further information.
Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand
Insert Length: 851 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 1.
Location/Qualifiers

FEATURES

source

1..37

/organism="Homo sapiens"

/db_xref="GDB:3828558"

/db_xref="taxon:9606"

/clone="IMAGE:184209"

/clone_lib="Soares adult brain N2b4HB55Y"

/sex="Male"

/dev_stage="55-year old"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: brain; Vector: pRT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'
TGTTCACATCTGAAGTGGAGCGCGCGCTTTTCTTTTCTTTT 3'),
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pRT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 53. Library constructed by Bento
Soares and M. Patricia Bernaldo. The adult brain RNA was
provided by Dr. Donald H. Gilden. Tissue was acquired
17-18 hours after death which occurred in consequence of a
ruptured aortic aneurysm. RNA was prepared from a pool of
tissues representing the following areas of the brain:
frontal, parietal, temporal and occipital cortex from the
left and right hemispheres, subcortical white matter,
basal ganglia, thalamus, cerebellum, midbrain, pons and
medulla."

BASE COUNT

6 a 8 c 11 g 10 t 2 others

ORIGIN

Query Match 49.6%; Score 11.4; DB 14; Length 37;
Best Local Similarity 65.2%; Pred. No. 4.8e+05;
Matches 15; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 AAATGGCTCCGAGCGGGAAC 23

DB 32 ACACCTGCTNANNGCGAGAAC 10

RESULT 14

AU254523 27 bp mRNA linear EST 25-APR-2002

LOCUS AU254523 3'-directed mouse cDNA library Mus musculus cDNA clone

DEFINITION BED0002375 3', mRNA sequence.

ACCESSION AU254523

VERSION AU254523.1 GI:20316384

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 27)

AUTHORS Kato, K. and Matoba, R.

TITLE Generation of expressed sequence tags from mouse brain

JOURNAL Unpublished (2002)

COMMENT Graduate School of Biological Sciences

Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara 630-0101, Japan
Tel: 81-743-72-5581
Fax: 81-743-72-5589
Email: kkatc@bs.ist-nara.ac.jp,
URL: http://love2.ist-nara.ac.jp/BED/index.html.

FEATURES

source

1..27

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="BED0002375"

/clone_lib="3'-directed mouse cDNA library"

/tissue_type="brain"

/note="Vector: pCEM-T-easy"

BASE COUNT 7 a 5 c 9 g 6 t

ORIGIN

Query Match 48.7%; Score 11.2; DB 9; Length 27;
Best Local Similarity 81.2%; Pred. No. 5.6e+05;
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 GCTCGAGCGCGGAAA 22

DB 12 GCTCTGAGCAGAGAAA 27

RESULT 15 27 bp mRNA linear EST 12-DEC-1995

LOCUS D18732 M03540.1794 Mouse 3'-directed Mus musculus domesticus cDNA clone

DEFINITION md0354 3', mRNA sequence.

ACCESSION D18732

VERSION D18732.1 GI:1100701

KEYWORDS EST.

SOURCE western European house mouse.

ORGANISM Mus musculus domesticus

REFERENCE 1 (bases 1 to 27)

AUTHORS Kawamoto, S., Okubo, K., Yoshii, J., Katsuki, M. and Matsubara, K.

TITLE Analysis of gene expression in mouse embryogenesis by 3'-directed

JOURNAL Unpublished (1995)

COMMENT Contact: Kawamoto, S., Okubo, K., Yoshii, J., Katsuki, M. and Matsubara

K., Institute for Cellular and Molecular Biology

Osaka University

3-1 Yamada-oka, Suita, Osaka 565, Japan.

Location/Qualifiers

1..27

/organism="Mus musculus domesticus"

/strain="C57BL/6J"

/db_xref="taxon:10092"

/clone="md0354"

/clone_lib="Mouse 3'-directed"

/tissue_type="decidual tissue (day 6.5-8.5 of gestation)"

BASE COUNT 7 a 5 c 9 g 6 t

ORIGIN

Query Match 48.7%; Score 11.2; DB 14; Length 27;
Best Local Similarity 81.2%; Pred. No. 5.6e+05;
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 GCTCGAGCGCGGAAA 22

DB 12 GCTCTGAGCAGAGAAA 27

Search completed: March 26, 2003, 23:17:58
Job time : 12237 secs

THIS PAGE BLANK (USPTO)